Brief Communication

Stochastic reinforcement benefits skill acquisition

Eran Dayan,[1,3] Bruno B. Averbeck,[2] Barry J. Richmond,[2] and Leonardo G. Cohen[1]

[1]*Human Cortical Physiology and Neurorehabilitation Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA;* [2]*Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA*

Learning complex skills is driven by reinforcement, which facilitates both online within-session gains and retention of the acquired skills. Yet, in ecologically relevant situations, skills are often acquired when mapping between actions and rewarding outcomes is unknown to the learning agent, resulting in reinforcement schedules of a stochastic nature. Here we trained subjects on a visuomotor learning task, comparing reinforcement schedules with higher, lower, or no stochasticity. Training under higher levels of stochastic reinforcement benefited skill acquisition, enhancing both online gains and long-term retention. These findings indicate that the enhancing effects of reinforcement on skill acquisition depend on reinforcement schedules.

[Supplemental material is available for this article.]

Learning new skills is driven by reinforcement, which can be either extrinsic, as in the form of monetary rewards (Wachter et al. 2009; Abe et al. 2011), or intrinsic (Shohamy 2011), as in a sense of fulfillment and pride. Normative models of valuation (Bell et al. 1988) view humans as reward-maximizing entities. Consequently, learning novel skills with reinforcement schedules where successful performance is continuously reinforced should maximally facilitate learning. Indeed, previous studies of reward-based motor skill acquisition utilized reinforcement schedules of this sort (Wachter et al. 2009; Abe et al. 2011). Yet, in ecologically valid settings complex skills are often acquired when mapping between actions and rewarding outcomes is not continuous and fixed but, rather, variable and unknown. This results in reinforcement schedules of a stochastic nature, and a state commonly referred to as uncertainty (Platt and Huettel 2008; Bach and Dolan 2012). For procedural skills, performance improvements can occur not only during training ("online"), but also between training sessions in periods where there is no active practice occurring ("offline"). These two forms of learning lead to formation of long-term memory (Doyon and Benali 2005; Dayan and Cohen 2011), and both appear to be affected by reinforcement (Wachter et al. 2009; Abe et al. 2011). Here, we studied how procedural learning would proceed under the incentive of stochastic reinforcement. We trained four groups of subjects ($n = 48$) on a sequential visuomotor task (Reis et al. 2009; Schambra et al. 2011), manipulating reinforcement schedules between groups. The task required subjects to move a cursor back and forth between a "home" position and five individually colored numbered targets (four gates and one thick line) by modulating pinch force applied onto a force transducer (Fig. 1A). Successful trials were ones where the sequence of movements (1-home, 2-home, 3-home, 4-home, 5) was performed accurately within a fixed amount of time (8 sec). Performance-based auditory feedback (a "beep" sound) was given whenever a target gate was passed through successfully. The experiment comprised three sessions (Fig. 1B). Following the first block of 20 trials, where baseline performance was assessed, reinforcement schedules were implemented for five blocks of training, where successful completion of each trial could result in visual reward feedback ("you win 0.6$"). Tests of skill levels were then administered with no reinforcement immediately after training as well as 24 h and 7 d post-training, to assess offline skill gains and long-term retention of skill, respectively. Training was carried out under four different reinforcement schedules (Fig. 1C), manipulated in a between subject design, varying reward probability to create four levels of stochastic reinforcement (with probability, P, at 0.25, 0.5, 0.75, 1). Under this experimental manipulation uncertainty as to whether a successful trial would get rewarded is maximal at $P = 0.5$ and decreases with higher and lower probabilities (Schultz et al. 2008), since these probabilities are associated with higher certainty pertaining to possible chances of being rewarded or unrewarded (Fig. 1D).

Skill acquisition was quantified per block using a skill measure combining movement time and error rates to capture shifts in the task's speed–accuracy trade-off function along training (Reis et al. 2009; Schambra et al. 2011; see Supplemental Methods). The two groups that trained with lower levels of stochasticity (0.25 and 0.75) showed indistinguishable performance across all four testing sessions ($F_{(3,66)} = 2.541$, ns), and therefore data from these two groups were collapsed into one group, henceforth the "low stochasticity" group ($n = 24$), and compared with a high stochasticity group (reward probability, $P = 0.5$, $n = 11$) and a fixed reward group (no stochasticity, reward probability, $P = 1.0$, $n = 11$). A repeated measures analysis of variance (ANOVA) revealed a significant interaction between the level of stochasticity and testing session ($F_{(6,129)} = 5.36$, $P < 0.0001$) (Fig. 2A). Specifically, all groups had comparable baseline performance levels prior to training ($F_{(2,43)} = 2.438$, ns). After training, the three training groups showed significantly different performance ($F_{(2,43)} = 4.794$, $P < 0.02$). Namely, the high stochasticity group showed significantly more skillful performance than the low stochasticity ($P = 0.023$, post hoc Fisher's least square differences test) and the fixed reward groups ($P = 0.004$). One day post-training, the three groups

[3]Corresponding author
E-mail dayane@ninds.nih.gov
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.032417.113.

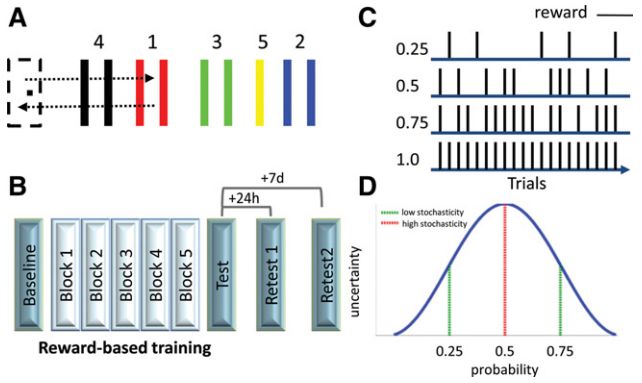

Figure 1. Task and design. (*A*) By varying the magnitude of pinch-force applied onto a force transducer, subjects moved a cursor back and forth via five numbered targets within a fixed period of time. (*B*) Experimental design. The experiment comprised three sessions, including a training session with reward feedback, followed by three tests of skill. (*C*) Reinforcement schedules. Four reinforcement schedules were tested, with reward feedback provided on 25%, 50%, 75%, or 100% of successful trials. (*D*) Reward uncertainty. Stochastic reinforcement was maximal and was associated with maximal levels of outcome uncertainty when reward probability was 0.5. With probabilities of 0.25 and 0.75, stochasticity and uncertainty were lower since the learning agents were operating with greater certainty pertaining to lower and higher chances of being rewarded, respectively.

maintained their levels of performance but skill levels did not differ ($F_{(2,43)} = 2.472$, ns). One week after training, most subjects in the high stochasticity (7/11) and the fixed reward (9/11) groups showed additional gains compared to their performance right after training. In the low stochasticity group, on the other hand, more subjects (13/24) showed evidence of forgetting (i.e., worse performance at 1 wk, compared to immediately after training). Overall, 1 wk after training, performance of the three training groups differed ($F_{(2,43)} = 8.71$, $P < 0.001$), whereby the high stochasticity group performed significantly better relative to the low stochasticity ($P < 0.0001$) and the fixed reward groups ($P < 0.001$). To evaluate the observed differences further, we also assessed the degree of online within-session gains (Fig. 2B), defined as the difference in skill between test and baseline (ΔTest−Baseline). Online gains were significantly different among the groups ($F_{(2,43)} = 4.668$, $P < 0.02$), with the group that trained under high stochasticity showing significantly larger gains relative to the two other groups ($P < 0.04$ and $P < 0.005$, when compared to the low stochasticity and fixed reward groups, respectively; see also Supplemental Results). Next we further assessed long-term retention of skill (Fig. 2C), defined as the difference between skill at 1 wk and performance immediately post-training (ΔRetest2−Test). This analysis again showed differences among the training groups ($F_{(2,43)} = 4.466$, $P < 0.02$), with the group training under high stochasticity showing better retention relative to the low stochasticity ($P < 0.005$) or to the fixed reward groups ($P < 0.05$).

The results reported here show that, contrary to what might be assumed from normative models of valuation (Bell et al. 1988), humans who learn new motor skills do not maximally benefit from training schedules where successful performance is continuously reinforced. Rather, training with high levels of stochastic reinforcement benefits skill learning more strongly, enhancing online within-session skill gains and further resulting in a stronger long-lasting memory trace of the acquired skill.

Sensorimotor control is carried out in the face of various sources of sensory, motor, and outcome uncertainties. Progress has been made recently in understanding how the brain controls movement facing inherent sensory and motor noise (Orban and Wolpert 2011). Yet, less is known about the behavioral consequences of outcome uncertainty and the possible strategies utilized to compensate for it, owing in part to lack of relevant empirical data (Bach and Dolan 2012). Earlier work established that removal of various forms of augmented extrinsic feedback about task success results in improved retention of skills (Schmidt et al. 1989; Winstein 1991). Augmented information feedback refers to the extrinsic feedback provided to the learner to support learning (Swinnen 1996). In the current experimental paradigm, visual and auditory performance feedback were provided in real time to allow subjects to perform the task accurately and did not differ across all training groups. Reinforcement, on the other hand, was provided probabilistically at the end of each trial, as an incentive for successful performance, independently from augmented information feedback. Important differences exist with respect to intermittent delivery of feedback and reinforcement (Swinnen 1996). Whereas removal of augmented information feedback affects retention but not learning (Schmidt et al. 1989; Winstein 1991), the current results documented enhancing effects of stochastic reinforcement on both learning and retention, further demonstrating the difference between how reinforcement and feedback guide skill acquisition.

Reinforcement-mediated motor learning has been shown in a variety of paradigms (Fischer and Born 2009; Wachter et al. 2009; Abe et al. 2011; Huang et al. 2011; Izawa and Shadmehr 2011; Shmuelof et al. 2012). Although it is still unresolved what constitutes a reward signal for the motor system (Wolpert et al. 2011), an emerging framework suggests that an underlying objective of voluntary movement is to achieve more valuable states (Shadmehr and Krakauer 2008). Along these lines, motor learning can be conceptualized as a process of optimization, where motor commands that minimize costs and maximize reward are shaped (Shadmehr and Krakauer 2008). To account for the findings reported here a mechanistic framework for reinforcement-mediated motor learning necessitates valuation systems that can integrate variables such as probability and magnitude to drive learning, of

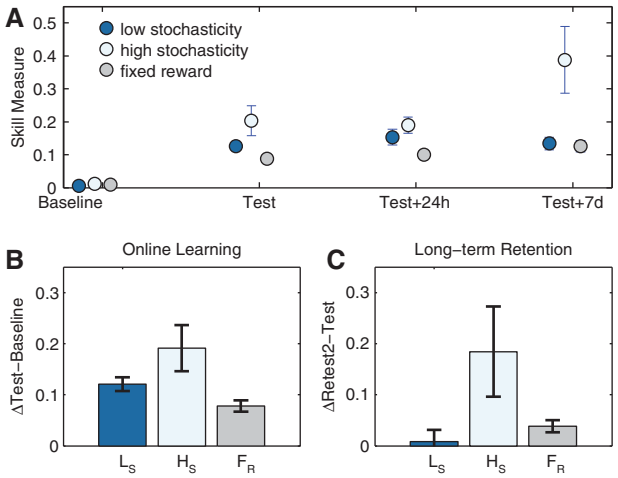

Figure 2. Training-related skill changes. (*A*) Changes of skill along training. Skill (a metric expressing shifts in the speed−accuracy trade-off function) at baseline, immediately after training (Test), 24 h later, and 7 d post-training. (*B*) Online learning. Online within-session gains were assessed by subtracting baseline skill scores from those measured immediately after training (test). (*C*) Long-term retention. To assess long-term retention, skill scores measured immediately after training were subtracted from scores measured 1 wk after training ended. Error bars depict SEM. (L_s) low stochasticity, (H_s) high stochasticity, (F_R) fixed reward.

the sort suggested by models of reinforcement learning (Kaelbling et al. 1996; Sutton and Barto 1998).

A range of information processing systems, both artificial and organic, appear to benefit from stochastic biological noise (McDonnell and Ward 2011), possibly by improving the system's overall signal-to-noise ratio. The beneficial effects of stochastic reinforcement may also stem from the influence these schedules and the uncertainty associated with them have on the saliency and the amount of attention directed toward reward-predicting cues (Pearce and Hall 1980; Dayan et al. 2000; Esber and Haselgrove 2011). Thus, training with stochastic reinforcement may render learning agents more susceptible to gain from training by allocation of more attentional and cognitive control resources during learning. The stochastic reinforcement schedules used in our experimental design may induce what has been referred to as "expected uncertainty," which results from a known unreliability of predictive relationships within a familiar environment (Yu and Dayan 2005). Previous modeling work linked expected uncertainty with faster learning (Yu and Dayan 2003), providing a possible mechanism for the within-session learning gains reported here. Other earlier findings based on animal models established that intermittent reinforcement schedules applied during operant conditioning are associated with greater resistance to extinction, expressed in increased response rate during this phase (e.g., lever presses), but no effects on acquisition (Humphreys 1939; Jenkins and Stanley 1950). It was also shown that the increased response rate varies monotonically with the thinning of the reinforcement schedule (Gallistel and Gibbon 2000). The current results challenge these earlier accounts showing effects during acquisition which do not vary monotonically with reward probability and are followed by a stronger memory trace. More generally, our results suggest that stochastic reinforcement also exerts effects on higher forms of learning such as the acquisition of complex novel visuomotor skills in humans, where successful task performance requires a skillful combination of speed and accuracy.

Acknowledgments

We thank G. Dold, G. Melvin, A. Harris, J. Hamann, and K. Zherdeva for technical help and N. Censor, B. Almog, H. Schambra, J. Reis, and Y. Niv for helpful suggestions. This work was supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke, National Institutes of Health.

References

Abe M, Schambra H, Wassermann EM, Luckenbaugh D, Schweighofer N, Cohen LG. 2011. Reward improves long-term retention of a motor memory through induction of offline memory gains. *Curr Biol* **21:** 557–562.

Bach DR, Dolan RJ. 2012. Knowing how much you don't know: A neural organization of uncertainty estimates. *Nat Rev Neurosci* **13:** 572–586.

Bell DE, Raiffa H, Tversky A. 1988. *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge University Press, Cambridge, UK.

Dayan E, Cohen LG. 2011. Neuroplasticity subserving motor skill learning. *Neuron* **72:** 443–454.

Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. *Nat Neurosci* **3 Suppl:** 1218–1223.

Doyon J, Benali H. 2005. Reorganization and plasticity in the adult brain during learning of motor skills. *Curr Opin Neurobiol* **15:** 161–167.

Esber GR, Haselgrove M. 2011. Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proc Biol Sci* **278:** 2553–2561.

Fischer S, Born J. 2009. Anticipated reward enhances offline learning during sleep. *J Exp Psychol Learn Mem Cogn* **35:** 1586–1593.

Gallistel CR, Gibbon J. 2000. Time, rate, and conditioning. *Psychol Rev* **107:** 289–344.

Huang VS, Haith A, Mazzoni P, Krakauer JW. 2011. Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models. *Neuron* **70:** 787–801.

Humphreys LC. 1939. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J Exp Psychol* **25:** 141–158.

Izawa J, Shadmehr R. 2011. Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput Biol* **7:** e1002012.

Jenkins WO, Stanley JC Jr. 1950. Partial reinforcement: A review and critique. *Psychol Bull* **47:** 193–234.

Kaelbling LP, Littman ML, Moore AW. 1996. Reinforcement learning: A survey. *J Artif Int Res* **4:** 237–285.

McDonnell MD, Ward LM. 2011. The benefits of noise in neural systems: Bridging theory and experiment. *Nat Rev Neurosci* **12:** 415–426.

Orban G, Wolpert DM. 2011. Representations of uncertainty in sensorimotor control. *Curr Opin Neurobiol* **21:** 629–635.

Pearce JM, Hall G. 1980. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* **87:** 532–552.

Platt ML, Huettel SA. 2008. Risky business: The neuroeconomics of decision making under uncertainty. *Nat Neurosci* **11:** 398–403.

Reis J, Schambra HM, Cohen LG, Buch ER, Fritsch B, Zarahn E, Celnik PA, Krakauer JW. 2009. Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proc Natl Acad Sci* **106:** 1590–1595.

Schambra HM, Abe M, Luckenbaugh DA, Reis J, Krakauer JW, Cohen LG. 2011. Probing for hemispheric specialization for motor skill learning: A transcranial direct current stimulation study. *J Neurophysiol* **106:** 652–661.

Schmidt RA, Young DE, Swinnen S, Shapiro DC. 1989. Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *J Exp Psychol Learn Mem Cogn* **15:** 352–359.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P. 2008. Explicit neural signals reflecting reward uncertainty. *Philos Trans R Soc Lond B Biol Sci* **363:** 3801–3811.

Shadmehr R, Krakauer JW. 2008. A computational neuroanatomy for motor control. *Exp Brain Res* **185:** 359–381.

Shmuelof L, Huang VS, Haith AM, Delnicki RJ, Mazzoni P, Krakauer JW. 2012. Overcoming motor "forgetting" through reinforcement of learned actions. *J Neurosci* **32:** 14617–14621.

Shohamy D. 2011. Learning and motivation in the human striatum. *Curr Opin Neurobiol* **21:** 408–414.

Sutton RS, Barto AG. 1998. *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA.

Swinnen SP. 1996. Information feedback for motor skill learning: A review. In *Advances in motor learning and control* (ed. Zelaznik HN), pp. 37–66. Human Kinetics, Champaign, IL.

Wachter T, Lungu OV, Liu T, Willingham DT, Ashe J. 2009. Differential effect of reward and punishment on procedural learning. *J Neurosci* **29:** 436–443.

Winstein CJ. 1991. Knowledge of results and motor learning—implications for physical therapy. *Phys Ther* **71:** 140–149.

Wolpert DM, Diedrichsen J, Flanagan JR. 2011. Principles of sensorimotor learning. *Nat Rev Neurosci* **12:** 739–751.

Yu AJ, Dayan P. 2003. Expected and unexpected uncertainty: ACh and NE in the neocortex. In *Advances in neural information processing systems* (ed. Becker S, et al.), **15:** 173–180. MIT Press, Cambridge, MA.

Yu AJ, Dayan P. 2005. Uncertainty, neuromodulation, and attention. *Neuron* **46:** 681–692.

Received July 11, 2013; accepted in revised form December 4, 2013.